SHENYAN ZENG

Tel: +86 15951991599 | **E-mail:** zengshenyan0308@163.com

EDUCATION

University of California, Santa Barbara
Bachelor of Science in Applied Mathematics — Sep 2021–Mar 2025

- **Relevant Courses:** Numerical Analysis, Linear Algebra, Real Analysis, Optimization, Graph and Network Theory, Probability and Statistics, Computer Science

RESEARCH EXPERIENCE

State Key Lab of CAD & CG, Zhejiang University — Hangzhou, China
Research Assistant (Advisor: Prof. Chunhua Shen) — Aug 2025–Present

- Co-authored a paper submitted to CVPR 2026 (3rd Author); contributed to the experimental infrastructure by building a unified evaluation pipeline for spatial reasoning across 10+ benchmarks.
- Benchmarked the proposed Qwen3-VL+DCRL model, confirming a substantial +43.0% F1 score improvement (27.5% → 70.5%) over the baseline and surpassing commercial models like GPT-4o on the ReasonMatch benchmark.
- Deployed reproducible pipelines for PointArena, OmniSpatial, and RoboSpatial on NAS clusters, implementing vLLM integration for Qwen2.5/3-VL with custom optimizations (e.g., FlashAttention fallback) to ensure robust inference.
- Developed advanced evaluation strategies including Circular Evaluation to mitigate positional bias in SAT tasks and Thread-safe Resume mechanisms, managing multi-GPU scheduling to maximize cluster efficiency.
- Debugged GIM Matching pipelines by analyzing RANSAC inlier distributions and fixing visualization distortions, providing critical insights that supported the paper's qualitative analysis.

INTERNSHIP EXPERIENCE

Wind Information Co., Ltd. — Nanjing, China
AI Algorithm Engineer Intern — May 2025–Aug 2025

- Led the fine-tuning of SmolVLM2 (98.8% OCR accuracy) and LayoutReader, successfully deploying them into the "DP 瞬析" WeChat Service to enable automated PDF-to-Excel conversion and real-time document digitization for end-users.
- Converted RT-DETR and PaddleOCR models to ONNX/TensorRT with FP16 quantization, reducing GPU memory usage by 25% and boosting inference speed by 3.2 × on A100 clusters.
- Designed a scalable cleaning and formatting workflow (Parquet/JSON) for a multi-modal dataset containing 1.25 million samples (handwriting, formulas, tables), establishing a standardized protocol for pre-training data readiness.
- Built automated diagnostics (IoU checks, failure visualization) for model outputs, reducing debugging cycles by ~40% and accelerating iteration efficiency for document-AI pipelines.

Wind Information Co., Ltd. — Nanjing, China
Intelligent Data Analytics Intern — Jul 2023–Sep 2023

- Cleaned & annotated 10k+ entries and standardized merges from DOCX/Excel to support model-training pipelines with higher dataset consistency
- Wrote Python automation to parse/aggregate model reports, cutting manual reporting time and improving reproducibility of evaluations

SKILLS

- **Programming Languages & Tools:** Python, SQL, Linux/Shell, Git, Docker
- **AI Frameworks:** PyTorch, HuggingFace (Transformers, PEFT), vLLM, PaddleOCR, OpenCV, Weights & Biases
- **Algorithms & Fields:** Multimodal LLMs (Qwen-VL, SmolVLM), Computer Vision (SfM, RANSAC, RT-DETR), NLP, RL
- **Benchmarking & Engineering:** Pipeline Design, Metric Standardization, Error Analysis (Circular Eval), ONNX, TensorRT, Parquet
- **Languages:** English (Fluent), Mandarin (Native)